

INTRODUCTION

A primary goal of testing is to collect information to use in making decisions. Depending on the kinds of decisions that need to be made, different types of information may be needed. This difference in the type of decision to be made forms the basis for two major types of tests - criterion-referenced tests (CRTs) and norm-referenced tests (NRTs). In criterion-referenced testing, the goal is usually to make a decision about whether or not an examinee can demonstrate mastery in an area of content and competencies. Oftentimes, the area of content and competencies being assessed is job-related; most certification and licensure exams are CRTs. In norm-referenced testing, the goal is usually to rank the entire set of examinees in order to make comparisons of their performances relative to one another. Many standardized educational tests are NRTs. The two types of tests differ in several additional important ways, including their comparison targets, the average item difficulty of the exams, the resulting examinee score distributions, and the types of scores typically reported.

DIFFERENCES BETWEEN CRTs AND NRTs

Comparison Targets

The most obvious difference between CRTs and NRTs is the comparison target, that is, what an examinee's performance is compared to. In CRTs the examinee's performance is compared to an external standard of competence or mastery. An examinee is classified as a master or non-master by either passing or failing the exam. In theory, there is no limit to the number of examinees who can succeed, or who can pass the exam. In NRTs the examinee's performance is typically compared to that of other examinees. On an NRT an examinee's opportunity for success is relative to the performance of the other individuals who take the test.

Average Item Difficulties

Another important difference between the two types of tests is the average item difficulty, or p-value, on the test forms. For CRTs, the average item p-value is likely to be fairly high since a majority of the examinees may be expected to demonstrate mastery, both on the individual items and the overall test. In NRTs, on the other hand, the average item p-value is likely to be quite a bit lower, as the items as a whole may be more difficult. Tests that have been designed in this way are better able to spread out the examinees'



scores and thus to provide a more reliable ranking of the examinees relative to one another.

Score Distributions

Another difference between the two types of tests is the shape of the examinees' score distributions. The average performance of examinees on a test is highly related to the average item p-value of that test. Thus, the typical difference between score distributions on CRTs and NRTs results from the differences in average item difficulty. For CRTs, where many or even most of the examinees do well on the test overall, a plot of the resulting score distribution will show most of the scores clustering near the high end of the score scale. With NRTs, a much broader spread of scores can be expected, with a few examinees earning very low scores, many earning medium scores, and a few examinees earning very high scores. This score distribution is sometimes called the "bell curve" or the "normal distribution." A normal distribution is not typical for CRT programs. If the score distribution for a CRT did look like the normal distribution, depending on the location of the passing score, it would probably suggest that only a small proportion of the examinees displayed mastery.

Reported Scores

CRTs and NRTs also differ in terms of the types of scores they usually report. For CRTs a simple classification decision is most commonly reported. This may be a classification of the examinee as master/non-master or pass/fail. For NRTs a score, rather than a classification, is more often reported; percentile ranks or scale scores are frequently used.

Summary

The goals of most certification and licensure exam programs are far more closely aligned to a CRT approach to test development than to an NRT approach. It is worth noting the characteristics of both, however, as many commonly available testing materials and analysis methods are specifically designed for NRTs. Being aware of these different characteristics enables you to interpret test materials and select analysis methods that are properly aligned with the goals of your exam program.

